Cholera toxin genes: nucleotide sequence, deletion analysis and vaccine development

John J. Mekalanos, Daryl J. Swartz & Gregory D. N. Pearson

Department of Microbiology and Molecular Genetics, Harvard Medical School, 25 Shattuck Street, Boston, Massachusetts 02115, USA

Nigel Harford, Francoise Groyne & Michel de Wilde

Department of Molecular Genetics, Smith-Kline-R.I.T., rue de l'Institut, 89, B-1330 Rixensart, Belgium

Nucleotide sequence and deletion analysis have been used to identify the regulatory and coding sequences comprising the cholera toxin operon (ctx). *Incorporation of defined* in vitro-*generated* ctx *deletion mutations into* Vibrio cholerae *by* in vivo *genetic recombination produced strains which have practical value in cholera vaccine development.*

MODERN history has recorded seven world pandemics of cholera, a diarrhoeal disease produced by the Gram-negative bacterium *Vibrio cholerae*[1]. Laboratory tests can distinguish two biotypes of *V. cholerae*, classical and El Tor, the latter being responsible for the most recent cholera pandemic. The diarrhoeal syndrome induced by colonization of the human small bowel by either biotype of *V. cholerae* is caused by the action of cholera toxin, a heat-labile enterotoxin secreted by the growing vibrios[2]. Diarrhoeal diseases affecting both humans and animals caused by some enterotoxinogenic strains of *Escherichia coli* are also induced by a heat-labile enterotoxin (LT) which is closely related to cholera toxin in structure and mode of action[1-3].

Cholera toxin is an 84,000-molecular weight (MW) protein composed of one A subunit (27,000 MW) and five B subunits (11,600 MW). The A subunit, although synthesized as a single polypeptide chain, is usually proteolytically nicked to form two disulphide-linked polypeptides, A1 (22,000 MW) and A2 (5,000 MW)[4,5]. The A1 polypeptide is an enzyme and promotes the activation of adenylate cyclase in target cells by catalysing the ADP-ribosylation of a GTP-binding regulatory component of the cyclase complex[6]. The resulting accumulation of cyclic AMP in the intestinal mucosa leads to the severe fluid loss characteristic of cholera. Each B subunit has a high binding affinity for the toxin's cell surface receptor, ganglioside GM_1 (ref. 7). Neutralizing antibodies raised against the holotoxin react mainly with the B subunits[1,2].

Much of the current interest in the genetics of cholera toxin has been promoted by the need to develop a more efficacious vaccine against this enterotoxic disease. Parenterally administered, killed whole-cell and toxoid vaccines have been shown to be largely ineffective in producing long-lasting immunity to cholera, presumably because they lack the ability to induce local immune responses in the intestine[8,9]. Since the natural disease is capable of inducing prolonged immunity[9,10], several investigators have proposed the use of attenuated, non-toxinogenic mutants of *V. cholerae* as live oral cholera vaccines[11-16]. While encouraging results in volunteer studies have been obtained with some of these strains, factors such as genetic instability or poor colonizing ability have contraindicated their use in the field[15-17].

The recent relaxation of US governmental guidelines prohibiting the molecular cloning of bacterial toxin genes has permitted the use of a powerful new approach to the analysis of cholera toxin gene structure and vaccine development. These studies have shown that like the *elt* genes, which encode *E. coli* LT[18], the genes for the A and B subunits of cholera toxin are arranged in a single transcriptional unit with the A cistron (*ctxA*) preceding the B cistron (*ctxB*)[19]. *V. cholerae* strains of the classical biotype contain a nontandem, chromosomal duplication of the *ctx* operon that is structurally identical in all strains.[53] In contrast, about 70% of El Tor strains have only a single copy of *ctx*, while the remaining strains have two or more *ctx* copies present on a tandemly repeated genetic element. This genetic duplication and amplification of the toxin operon may be related to the instability observed in some of the earlier *V. cholerae* toxin mutants[13-16].

In this article, we report the entire nucleotide sequence of one *ctx* operon together with partial sequences containing the *ctx* promoter regions of five other cloned *ctx* copies. Deletion analysis has allowed the identification of toxin transcriptional and translational regulatory sequences. An *in vitro*-constructed, internal deletion in *ctxA* was recombined *in vivo* into both *ctxA* gene copies of *V. cholerae* strain Ogawa 395. Since this genetic recombinant still produces the immunogenic B subunit of the toxin, it should have practical value in cholera vaccine development.

Molecular cloning of *ctx* operon copies

A total of six *ctx* copies were cloned from four *V. cholerae* strains. These include both *ctx* copies from strain 569B, both copies from strain RV79, one of two copies from strain E7946 and the single *ctx* copy of strain 2125. With the exception of the classical strain 569B, all these strains are El Tor in biotype. The *ctx* copies were cloned as various *V. cholerae* restriction fragments which hybridized with ^{32}P-labelled *elt* or *ctx* probes as described in the legend to Fig. 1.

The restriction sites for several endonucleases were located on these cloned inserts, and the resulting maps were aligned at the conserved *Xba*I site previously determined to lie early in the A cistron[19], (Fig. 1*A*). Other conserved restriction sites for *Nru*I, *Pst*I, *Ava*I and *Bgl*II were also found preceding *ctxA* on these various inserts (Fig. 1). Additional restriction mapping and hybridization analysis has indicated that the 5 kilobase pairs (kbp) of DNA directly preceding the toxin structural genes is the same for all cloned *ctx* copies thus far examined[53]. Since the larger chromosomal sequence environment flanking these different *ctx* copies appears to vary as determined by Southern blot hybridization, we have proposed that the conserved DNA immediately upstream of *ctx* is part of a genetic element responsible for toxin operon duplication and transposition events. From a practical point of view, the conserved 5′ flanking sequences associated with *ctx* provided part of the necessary homology for efficient *in vivo* recombination of *in vitro* constructed deletion mutations into multiple copies of the toxin operon (see below).

Nucleotide sequence of *ctx*

The strategy used to determine the complete nucleotide sequence of the single *ctx* copy of strain 2,125 is shown below the restriction map of the insert cloned on pRIT10824 (Fig. 1*B*). The 2,020 nucleotides determined are shown in Fig. 2. Comparison of the 2,125 nucleotide sequence with both the *elt* nucleotide sequence[20,21] and known amino acid sequences of

0028-0836/83/490551—07$01.00

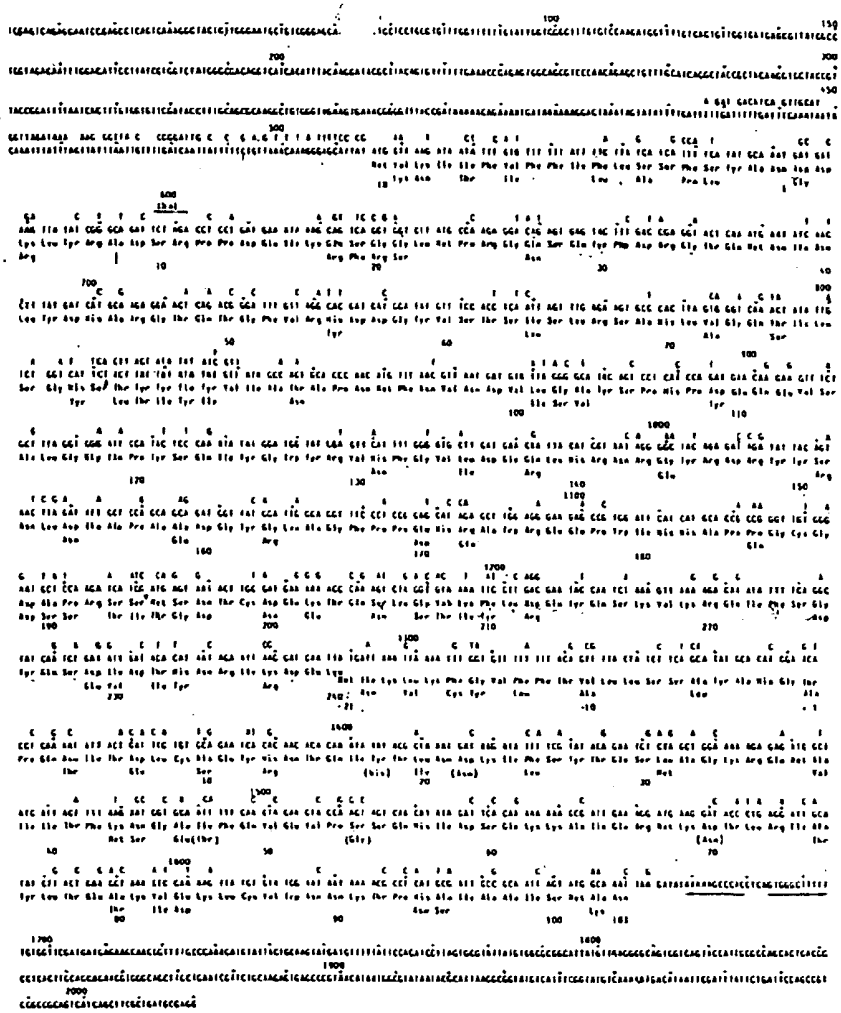

Fig. 2 DNA sequence of the *V. cholerae* toxin operon from strain 2125. The antisense strand is shown from 5' to 3'. From nucleotide 427 to 1,663, the sequence is compared with the published sequence of LT genes[20,21]; *elt* nucleotides are shown above the sequence only where they differ from the *ctx* sequence except between nucleotides 810 and 830, where deletion of a T (arrowed) in the *elt* sequence creates a frameshift which is corrected by insertion of a C (arrowed) 16 bp downstream (see text). Analogous events have previously been seen after pseudoreversion of frameshift mutations[43] but to our knowledge, this is the first naturally occurring case described. The deduced amino acid sequence of *ctxA* (nucleotide 516 to 1,289) and *ctxB* (nucleotide 1,289 to 1,660) is shown and compared with that of LT. Amino acids that differ in LT are shown below the cholera toxin amino acid sequence. In addition, for the mature B subunit sequence, differences from the published amino acid sequence of B subunit purified from strain 569B[16,17] are shown in brackets. Two of these differences (amino acids 47 and 54) are also found in LT. The cleavage site between the A1 and A2 polypeptides is indicated by an arrow (amino acids 194–195). Note also the overlap of *ctxA* and *ctxB* cistrons (nucleotides 1,289–1,292). Sequence exhibiting dyad symmetry and potentially involved in transcription termination is indicated with divergent arrows. Features of the sequence immediately upstream of the *ctxA* gene are detailed in Fig. 3.

sequence (nucleotides 1,277–1,282, Fig. 2) of the *ctxA* cistron. The first two nucleotides of the *ctxA* translation termination signal TGA are the last two nucleotides of the *ctxB* translation initiation triplet ATG. This particular overlapping arrangement is also found several times in phage λ operons[30] and may be involved in translational coupling[31] of the *ctxA* and *ctxB* genes. However, evidence presented below suggests that this is not the case with the *ctx* operon. Where documented, translational coupling is observed between cistrons whose gene products interact in a one to one stoichiometry[31], and in contrast, the cholera toxin molecule is composed of one A subunit and five B subunits. Moreover, *E. coli* produces stoichiometrically 7 times more cholera toxin B subunit than A subunit (data not shown). Fusion of the *ctxB* gene to various *E. coli* promoters allows high expression of *ctxB* in the absence of *ctxA* translational initiation signals. These data suggest that translation of *ctxB* relies primarily on independent initiations promoted by its own ribosome binding site.

Another experiment supports this conclusion. Our DNA sequencing analysis identified two *Nde*I sites at positions 561 and 1,337 within the *ctxA* and *ctxB* genes, respectively. The positions of these sites relative to the reading frames of *ctxA* and *ctxB* allowed us to construct a *ctxA* deletion which codes for an in-frame fusion of amino acid 17 of the A subunit signal sequence to amino acid 19 of the B signal and thus maintains the normal processing site of the B signal sequence (residue 21). This genetic fusion makes B subunit expression dependent on the efficiency of the A cistron translation initiation sequences, provided the hybrid signal sequence is processed at normal efficiency. *Nde*I digestion of plasmid pGP3 followed by ligation produced such a fusion between these two sites and gave plasmid pJM3.1. Plasmid pJM3.1 produced 0.056 μg ml^{-1} of B subunit in *E. coli* MS371 while pGP3 produced 0.50 μg ml^{-1}. These data suggest that the *ctxB* ribosome binding site is about ninefold more efficient than the *ctxA* site.

Toxin promoter regions

We determined approximately 200 base pairs of sequence upstream of the *Xba*I sites for each of the other five additional cloned copies of the *ctxA* gene, cloned on plasmids pGP3, pGP4, pGP5, pGP6 and JM17. Comparison of these sequences with the corresponding region of the *ctxA* gene derived from strain 2,125 indicated a perfect conservation of sequence between these copies from nucleotides 413 to 590 with one notable exception. The sequence TTTTGAT comprising nucleotides 419–425, 426–432 and 433–439 of the 2,125 sequence was found tandemly repeated 3–8 times preceding different *ctxA* gene copies (Fig. 3). Figure 3 shows part of a sequencing gel autoradiograph that spans DNA carrying eight of these tandem repeats in the region adjacent to the *ctxA* gene of pJM17.

To determine the position of the toxin operon promoter with respect to these repeated sequences, we used nuclease *Bal*31